一种基于改进分层置信规则库的社交账户可信度评估方法

吴 菲[†]，王 维
(长春工业大学 数字媒体教研室，长春 130012)

摘 要：社交账户可信度评估是确保网络社交生态良性发展的重要环节。针对社交账户可信度评估指标多维、数据信息不确定性多样等问题，提出了一种基于改进分层置信规则库的可信度评估方法。首先从账户属性、交际属性和内容属性三个角度分析了可信度评估各指标之间的相互关系，并依此构建了置信规则库的分层结构。其次，在信息转换函数中引入了自适应系数以更好描述和处理指标间的特性差异。最后，为了弥补专家知识局限性带来的模型误差，采用带有投影算子的协方差矩阵自适应进化策略对自适应系数和模型参数进行了优化。以新浪微博账户作为实验对象，结果表明该方法能够在数据样本有限的情况下获得更高的可信度评估精度。

关键词：置信规则库；社交账户；可信度评估
中图分类号：TP3 doi: 10.19734/j.issn.1001-3695.2022.01.0047

Credibility evaluation method for social accounts via improved hierarchical belief rule base

Wu Fei[†], Wang Wei
(*Dept. of digital media*, *Changchun University of Technology*, *Changchun Jilin* 130012, *China*)

Abstract: Social account credibility evaluation is an important link to ensure the benign development of network social ecology. Aiming at the problems of multi-dimensional credibility evaluation indexes and various data information uncertainty, this paper proposed a credibility evaluation method based on the improved hierarchical belief rule base. Firstly, this paper constructed a hierarchical structure by analyzing the relationship between the indicators of credibility evaluation from three perspectives: account attribute, communication attribute, and content attribute. Secondly, this paper introduced an adaptive coefficient into the information transformation function to better deal with the characteristic differences between indicators. Finally, to make up for the model error caused by the limitation of expert knowledge, this paper used the covariance matrix adaption evolution strategy with projection to optimize the adaptive coefficients and model parameters. Taking Sina Weibo account as the experimental object, the results show that this method can obtain higher accuracy when the data samples are limited.

Key words: belief rule base; social accounts; credibility evaluation

0 引言

随着互联网技术的不断发展，网络社交媒体逐渐成为了人们发布、传播、获取信息的主要方式。社交媒体为人们带来生活便利的同时，其开放共享的信息传播机制也逐渐成为了我国意识形态、信息安全、防疫防控等领域的风险隐患[1,2]。尤其是，在国际局势不断恶化、新冠肺炎等全球公共安全事件频繁发生的大背景下，境外极端份子借助社交媒体散布谣言和钓鱼链接等违法信息，欲达到实施网络诈骗、窃取国家机密甚至颠覆政权的目的。这些行为严重威胁了网络社交生态的良性发展，造成了社会舆论引导混乱，影响了社会安定。

社交账户是媒体信息发布的源头。准确判断账户是否可信，有利于相关部门采取合适的手段来对危害信息进行管控。现有的研究通常采用包括机器学习、统计分析等在内的多种建模方法对社交账户的状态进行判断[3,4]。王峥等人提出了一种特征加权贝叶斯神经网络模型，并将之应用于微博账号的异常检测中[5]。但是该模型依赖于高质量的训练数据样本。胡学韬等人基于粗糙集理论设计了社交账户信任度模型，通过该模型可以将社交账户的状态区分为正常和异常[1]。路金泉等人基于贝叶斯算法和层次分析法提出了一种账户可信度评估方法，该方法将社交账户评估为可信与不可信两级[2]。这两种方法能够分别有效处理模糊不确定性和概率不确定性，但对账户可信度的量化分析不够。刘亚尚等人的研究中，将社交账户的状态分为正常账户、被入侵账户和僵尸账户三类，采用所提的并行支持向量机算法实现了账户状态的识别[4]。该方法对训练样本数量的要求不高，但同样在不确定性描述和可信度量化方面存在短板。基于 D-S 证据理论的可信度评估方法能够有效融合专家主观判断和有限的客观数据，其采用置信辨识框架来描述信息的不确定性。但该评估方法在处理冲突证据方面存在不足[6,7]。

置信规则库(Belief Rule Base, BRB)是英国曼彻斯特大学杨剑波教授在 D-S 证据理论、模糊理论和 IF-THEN 规则的基础上发展而来的一种基于半定量信息的评估方法[6,7]。该方法通过在传统 IF-THEN 规则中引入置信框架来量化描述各类不确定性，通过在 D-S 证据理论中引入证据权重形成证据推理算法(Evidential Reasoning, ER)来处理冲突证据。BRB 能够有效融合专家判断和样本数据进行建模，降低了对高质量数据集的依赖[6,7]。目前，BRB 已经广泛应用于复杂工程系统的健康状态评估、性能评估和大型工业结构的安全性评估中[7,8]。

账户可信度的评估涉及多维指标，如用户关注数、评论率等，每个指标具有不同的特性。因此，要充分结合指标含义构建多层评估指标体系。指标体系构建完毕后，可采用分层 BRB 实现对多维多层指标的评估。然而，在现有分层 BRB 模型中，不同指标常常共用一种信息转换方法，且这些方法

收稿日期：2022-01-18；**修回日期**：2022-04-01

作者简介：吴菲(1983-)，女(通信作者)，吉林长春人，讲师，硕士，主要研究方向为计算机技术(83950185@qq.com)；王维(1982-)，女，吉林吉林人，讲师，硕士，主要研究方向为计算机技术.

不具备自适应性，不能较好反映个指标的特点。鉴于此，本文提出了基于改进分层 BRB 的账户可信度评估方法(Improved Hierarchical Belief Rule Base, IHBRB)，主要贡献总结如下：

1)提出了账户可信度评估指标，并依此构建了分层 BRB 评估模型；

2)提出了一种自适应指标信息转换方法，并通过智能优化算法实现了相关参数的自适应调整。

1　研究思路

本文提出的基于 IHBRB 的应用于社交账户可信度评估的主要思路如图 1 所示。首先，进行可信度评估指标的选取，并结合指标构建多层评估指标体系；然后，针对指标体系构建 IHBRB 模型，提出自适应信息转换方法；最后，构建优化模型，将 IHBRB 的置信度、属性权重、规则权重、参考值和信息转换方法参数等作为待优化参数进行自适应调整。优化后的评估模型即可用于社交账户可信度评估。

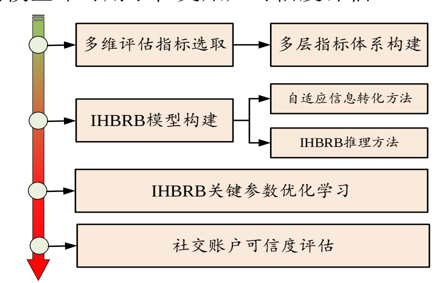


图 1　基于 IHBRB 的社交账户可信度评估研究思路

Fig. 1　Research idea of the social account credibility evaluation based on IHBRB

2　可信度评估指标体系构建

结合现有研究，本着“指标信息获取成本低”的原则，从多个角度选取了能够反映账户状态的指标[1, 9~11]。

账户年限，记为 C_1。该指标能够反映了账户使用的历史。在网络社交中，极端份子为了短时间内制造高强度的舆论压力，通常需要短时间内创建较大规模数量的僵尸账号。从这个角度出发，账户年限越短，其是僵尸账号的可能性越大，其可信度也就越低。

认证情况，记为 C_2。该项指标涉及身份认证、兴趣认证、问答认证、会员认证等，其能够反映账户信息的完备情况。在社交媒体中，非认证账户更加普遍，进行违法信息生产和传播的成本也更小。因此，可认为非认证账户的可信度要低于认证账户。

粉丝数和被转赞评数，分别记为 C_3 和 C_4。这两项指标主要反映账户的影响力。通常，粉丝数多的账户其可信度更高，相应的被转赞评数也会越多。但是部分异常账户发布谣言后，在僵尸账户的操纵下，其被转赞评数量也可能会出现增多的情况。因此，对于粉丝数少且被转赞评数多的账户，其可信度可能处于较低水平。

信息原创率，记为 C_5。该项指标可采用如下公式计算：

$$C_5=\frac{r_o}{r_a},0\le C_5\le 1 \tag{1}$$

其中，r_o 表示账户发表的原创信息条数，r_a 表示账户所发表信息的总条数；信息原创率能够从一定程度上反映用户自身的活跃度。信息原创率越高的账户，是僵尸账户或被入侵账户的可能性越低。

信息存疑率，记为 C_6。无论是原创信息还是转发信息，当信息内容含有异常链接、杂乱表情以及无含义文字时，可认为该信息存在违规嫌疑。信息存疑率可采用如下公式计算：

$$C_6=\frac{r_s}{r_a},0\le C_6\le 1 \tag{2}$$

其中，r_s 表示账户发表的存疑信息条数。所发布信息的存疑率越高，账户的可信度就越低。

上述六个指标中，账户年限 C_1 和认证情况 C_2、粉丝数 C_3 和被转赞评数 C_4、信息原创率 C_5 和存疑率 C_6 分别从账户属性、交际属性和内容属性三个角度反映了账户的可信度。基于这些角度反映的可信度便可进一步对账户的整体可信度进行评估。

因此，可构建如图 2 所示的账户可信度评估指标体系。其中，用 B 表示社交账户的可信度，用 B_1、B_2 和 B_3 分别表示账户属性可信度、交际属性可信度与内容属性可信度。

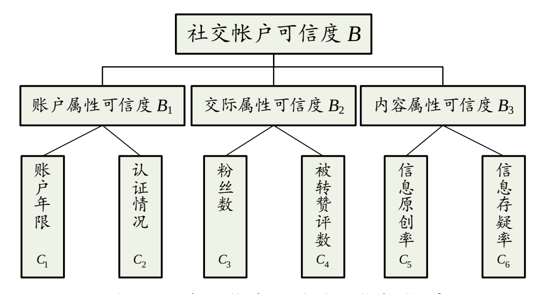


图 2　账户可信度评估分层指标体系

Fig. 2　Hierarchical index system for account credibility evaluation

3　基于 IHBRB 的可信度评估模型构建

3.1　基于 IHBRB 的可信度评估框架

结合分层指标体系，可构建如图 3 所示的基于 IHBRB 的可信度评估模型框架。其中，BRB-1 子模型、BRB-2 子模型和 BRB-3 子模型分别用于建立三种属性可信度与各指标之间的关系，而 BRB-3 子模型则用于建立三种属性可信度与账户可信度之间的非线性关系。IHBRB 模型采用由底至上的逐层推理模型得到最后的结果。

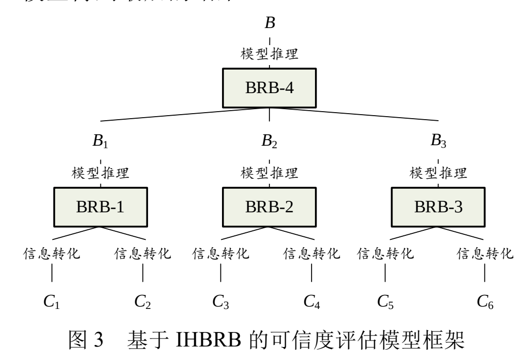


图 3　基于 IHBRB 的可信度评估模型框架

Fig. 3　Credibility evaluation model framework based on IHBRB

IHBRB 模型中的第 h 个子模型记为 BRB-h，它有一系列置信规则组成，其中第 k 条规则描述为

$$\begin{aligned}&R_k^h:\text{if } X_1^h \text{ is } A_1^{k,h}\wedge\dots\wedge X_M^h \text{ is } A_M^{k,h},\text{ then}\{(D_1^h,\beta_{1,k}^h),\dots(D_N^h,\beta_{N,k}^h)\},\\&\quad\text{with rule weight } \theta_k^h \text{ and attribute weights } \delta_i^h\ (i=1,\dots,M)\end{aligned} \tag{3}$$

其中，X_i^h 表示模型的第 i 个输入指标，例如 BRB-1 中的 C_1 和 C_2；$A_i^{k,h}(k=1,\dots,L)$ 表示第 k 条规则中 X_i^h 的参考等级，L 为规则的总条数，M 表示输入指标的数量；$\wedge$ 表示逻辑关系“与”；θ_k^h 表示第 k 条规则的规则权重；同理，δ_i^h 表示 X_i^h 的权重，称为属性权重；$\{(D_1^h,\beta_{1,k}^h),\dots(D_N^h,\beta_{N,k}^h)\}$ 表示第 k 条规则结论部分中各个不同可信度等级的置信分布。其中，$0\le\beta_{n,k}^h\le 1,(n=1,2,\dots,N)$ 表示第 k 条规则中对于第 n 个可信度等级 D_n^h 的支持度，也称对 D_n^h 的置信度。

输入信息转换是 IHBRB 模型的关键步骤，其主要目的是采用合适的转换方法将多种形式的输入信息统一至置信框架下。本文在传统信息转换方法的基础上提出了一种新的自适应信息转换方法。

3.2　自适应信息转换方法

IHBRB 模型输入指标 X_i^h 转换为置信分布形式的过程可用如下公式表示：

$$f(X_i^h)=\{(A_{i,j}^h,\alpha_j^{i,h}),j=1,2\cdots J,\ i=1,2\cdots M\} \tag{4}$$

其中，$\alpha_j^{i,h}$ 表示第 i 输入指标相对于第 j 个参考等级的匹配度，J 表示参考等级的个数，$A_{i,j}^h$ 表示属性参考值，$f(\bullet)$ 表示转换函数。在现有研究中，传统 BRB 模型的定量输入通常采用式(5)所示转换方法：

$$\alpha_j^{i,h}=f(x_i)=\begin{cases}\dfrac{A_{i,(j'+1)}^h-x_i}{A_{i,(j'+1)}^h-A_{i,j'}^h} & j'=j\ \text{if}\ A_{i,j'}^h\le x_i\le A_{i,(j'+1)}^h\\ 1-\dfrac{A_{i,(j'+1)}^h-x_i}{A_{i,(j'+1)}^h-A_{i,j'}^h} & j=j'+1\\ 0 & j=1,2,\ldots J_i,\ j\ne j',\ j\ne j'+1\end{cases} \tag{5}$$

其中，x_i 表示第 i 指标的输入值。可以看出，上述转换方法是线性的，难以准确描述输入与置信分布间的非线性关系。

鉴于此，对式(5)进行一般化构造，增强其非线性描述能力。构造的转换方法如下：

$$\alpha_j^{i,h}=f'(x_i)=\begin{cases}\left[\dfrac{A_{i,(j'+1)}^h-x_i}{A_{i,(j'+1)}^h-A_{i,j'}^h}\right]^s & j'=j\ \text{if}\ A_{i,j'}^h\le x_i\le A_{i,(j'+1)}^h\\ \left[1-\dfrac{A_{i,(j'+1)}^h-x_i}{A_{i,(j'+1)}^h-A_{i,j'}^h}\right]^s & j=j'+1\\ 0 & j=1,2,\ldots J_i,\ j\ne j',\ j\ne j'+1\end{cases} \tag{6}$$

其中，$f'(\bullet)$ 表示改进后的转换函数；$A_{i,j'}^h\le A_{i,(j'+1)}^h$ 分别表示相邻两个参考值；$s,s>0$ 为自适应系数，其决定着 $f'(\bullet)$ 的非线性能力，其可由专家给定，也可通过优化的方式得到。

为了说明这一点，假设 $A_{i,j'}^h\le x_i\le A_{i,(j'+1)}^h,A_{i,j'}^h=1,A_{i,(j'+1)}^h=10$，且 $g(x_i)=\left[{(A_{i,(j'+1)}^h-x_i)}\big/{(A_{i,(j'+1)}^h-A_{i,j'}^h)}\right]^s$，当 s 取不同值时，$g(x_i)$ 的输出如图 4 所示。可以看出，当 $0<s<1$ 时，$g(x_i)$ 为凹函数；当 $1<s$ 时，$g(x_i)$ 为凸函数；特别地，当 s=1 时，式(6)退化为式(5)，其描述的是线性关系。

在实践中，针对不同指标的转换方法赋予不同的自适应系数，即可实现更加精准和有效的信息转换。

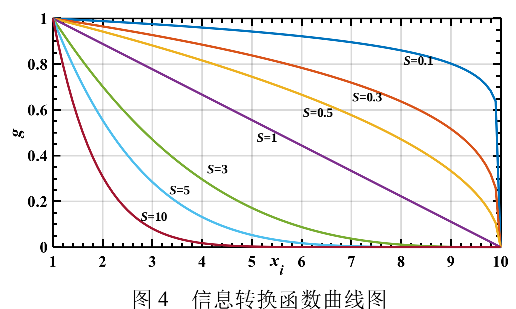


图 4　信息转换函数曲线图

Fig. 4　The curve of the information transformation function

3.3　可信度评估模型推理方法

IHBRB 模型中，各子模型的输入信息在转换为置信分布后，需要进行规则激活和融合，其输出结果作为下一层子模型的输入，依次完成逐层推理。不同子模型间的推理模式相同，以第 BRB-h 为例，其推理的主要步骤如下：

Step1(规则的激活)：在获取匹配度后，需要结合属性权重和规则权重来计算相应规则的激活权重。规则激活权重用以表示输入信息对规则的激活程度，其计算方法为

$$w_k^h=\frac{\theta_k^h\prod_{i=1}^{M}(\alpha_i^{k,h})^{\bar{\delta}_i^h}}{\sum_{l=1}^{L}\theta_l^h\prod_{i=1}^{M}(\alpha_i^{l,h})^{\bar{\delta}_i^h}}\ ,\ \bar{\delta}_i^h=\frac{\delta_i^h}{\max\limits_{i=1,\ldots,M}\delta_i^h} \tag{7}$$

其中，w_k^h 表示第 k 条规则的激活权重，当 $0<w_k^h\le 1$ 时则认为相应规则已被激活；θ_k^h 表示规则权重；$\alpha_i^{k,h}$ 表示该规则中第 i 个输入相对其参考值的匹配度；δ_i^h 表示属性权重；$\bar{\delta}_i^h$ 表示归一化后的相对属性权重。

Step2(规则的融合)：对于激活的规则，可使用证据推理(evidential reasoning，ER)算法将规则融合，ER 解析算法表达式为

$$\hat{\beta}_n^h=\frac{\mu\left[\prod_{k=1}^{L}\left(w_k^h\beta_{n,k}^h+1-w_k^h\sum_{i=1}^{N}\beta_{i,k}^h\right)-\prod_{k=1}^{L}\left(1-w_k^h\sum_{i=1}^{N}\beta_{i,k}^h\right)\right]}{1-\mu\left[\prod_{k=1}^{L}(1-w_k^h)\right]} \tag{8}$$

$$\mu=\left[\sum_{n=1}^{N}\prod_{k=1}^{L}\left(w_k^h\beta_{n,k}^h+1-w_k^h\sum_{i=1}^{N}\beta_{i,k}^h\right)-(N-1)\prod_{k=1}^{L}\left(1-w_k^h\sum_{i=1}^{N}\beta_{i,k}^h\right)\right]^{-1} \tag{9}$$

其中，$\beta_{n,k}^h$ 表示第 k 条规则对第 n 个参考等级 D_n 的置信度；$\hat{\beta}_n^h$ 表示输出中第 n 个参考等级 D_n 的置信度，且满足 $0\le\hat{\beta}_n^h\le 1,\sum_{n=1}^{N}\hat{\beta}_n^h\le 1$。

Step3(评估结果的输出)：推理后得到的评估结果可表示为如下所示的置信分布形式：

$$S(x_i)=\{(D_n,\hat{\beta}_n^h)\},n=1,2,\ldots,N \tag{10}$$

其中，$S(\bullet)$ 表示输出函数；上述结果以效用的形式输出可表示为

$$\hat{y}(x_i)=\mu(S(x_i))=\sum_{n=1}^{N}\mu(D_n)\hat{\beta}_n^h \tag{11}$$

其中 $\mu(D_n)$ 表示参考等级 D_n 的效用，即其参考值；$\mu(S(x_i))$ 为输出结果的效用。

在 IHBRB 中，$s(x_i)$ 作为下一层子模型的输入信息进行推导，直到生成最终结果。

4　基于 IHBRB 的可信度评估模型的自适应优化

对于社交账户的可信度评估问题，现有理论方法难以建立起精确的机理模型，但用户和专家在长期使用过程中能够积累一定的经验，通过网页爬虫等技术手段也可以获取一些数据样本。本文所提方法能够综合利用上述信息，即 IHBRB 模型的初始参数可由领域专家结合经验给定，后通过优化算法和数据样本对初始参数进行调整，以弥补专家知识局限性造成的模型误差[7]。

现有关于 BRB 模型优化的研究主要分为两类[12,13]：一类是在一定约束下调整待优化参数，使模型输出与实际系统输出之间的误差最小；另一类是在优化目标中引入结构参数，在提高模型建模精度的同时降低模型的复杂度。然而，这两类方法未在优化时考虑转换函数特性对建模性能的影响。鉴于此，在现有优化目标函数中引入转换函数的自适应系数作为待优化参数，新的优化目标函数及其约束条件可构建如下：

$$\begin{aligned}&\min\varphi(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{s})\\&\boldsymbol{\theta}=(\theta_1^1,\ldots,\theta_k^h,\ldots,\theta_L^H),\boldsymbol{\beta}=(\beta_{1,1}^1,\ldots,\beta_{n,k}^h,\ldots,\beta_{N,L}^H)\\&\boldsymbol{\delta}=(\delta_1^1,\ldots,\delta_i^h,\ldots,\delta_M^H),\boldsymbol{s}=(s_1,\ldots,s_i,\ldots,s_M)\\&\text{s.t.}\\&0\le\theta_k^h\le 1,k=1,2,\ldots L;h=1,2,\ldots,H\\&0\le\delta_i^h\le 1,\ \ i=1,2,\ldots M\\&0\le\beta_{n,k}^h\le 1,\ \ n=1,2,\ldots N,\sum_{n=1}^{N}\beta_{n,k}^h=1\\&0<s_i,i=1,2,\ldots M\end{aligned} \tag{12}$$

其中，$\boldsymbol{\theta}$，$\boldsymbol{\beta}$，$\boldsymbol{\delta}$ 和 s 分别为 IHBRB 中所有规则权重、置信度、属性权重和转换函数自适应系数构成的参数向量；H 为 IHBRB 中子模型的个数；$\varphi(\bullet)$ 为损失函数，用以描述模型输出与实际系统输出的区别。在本文中，损失函数用均方误差来计算，其具体描述如下

$$\varphi(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{s})=\frac{1}{T}\sum_{t=1}^{T}(y_i-\hat{y}_i)^2 \tag{13}$$

其中，T 为输出的数量；y_i 和 $\hat{y}_i$ 分别表示实际输出和模型输出。

现有研究已证明 BRB 模型是由多个复合函数组成的非

线性非凸模型，这为其参数优化带来了挑战[14]。为了提高建模精度，许多优化算法被应用于 BRB 的优化，例如差分进化算法(Differential Evolution，DE)、粒子群优化算法(Particle Swarm Optimization，PSO)和带有投影算子的协方差矩阵自适应进化策略(Covariance Matrix Adaption Evolution Strategy with Projection, P-CMA-ES)[5~18]。与 DE 算法和 PSO 算法相比，P-CMA-ES 算法能够以专家确定的初始解为中心，以多维正态分布形式进行新解的生成，这有利于在模型优化过程中充分结合专家的初始判断[15]。同时，P-CMA-ES 算法在高维非线性优化方面表现优越，能够在寻优过程中快速收敛至全局最优点[15]。P-CMA-ES 的基本流程如图 5 所示。

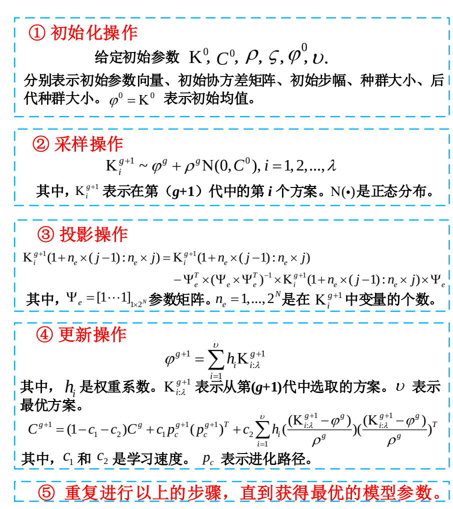


图 5　P-CMA-ES 算法的主要流程

Fig. 5　Main flow of P-CMA-ES algorithm

5　案例研究

5.1　案例背景

新浪微博是国内典型的热门社交媒体，是谣言等危害信息发布的重要场所。在其长期的运营中，积累了大量的账户资料。本文以新浪微博中的社交账户作为研究对象，通过开放数据接口和通用爬虫技术对账户注册年限 10 年内的账户资料进行了搜集，获取了 100 个账户的基本信息。通过人工分析这些账户的交际行为、内容发布等诸多基本特征，将其分为了“完全不可信”、“部分可信”、“基本可信”三类。

社交账户可信度评估指标对应的数据如图 6 所示。可以看出，账户的多种属性与其年限均有一定关联性。但由于用户存在使用习惯、受教育水平以及生活地域等多方面差异，这些数据难以直观反映账户可信度，因此必须采用一定的建模方法实现可信度评估。本文将采用所提基于 IHBRB 的社交账户可信度评估方法开展实例研究。

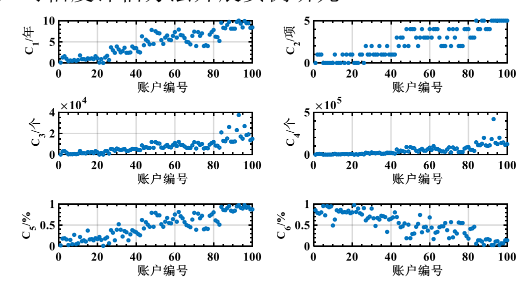


图 6　账户可信度评估指标数据

Fig. 6　Data of account credibility evaluation indicators

5.2　案例分析与模型构建

使用 IHBRB 进行账户可信度评估时，首先需要选择评估指标及可信度的参考等级。

对于账户属性可信度、交际属性可信度和内容属性可信度，其是社交账户可信度的局部估计，因此其参考等级可与社交账户可信度设置的一致，即“完全不可信(U)”、“部分可信(P)”和“基本可信(F)”。对应的参考值可分别取 0, 0.5 和 1。

对于评估指标，其参考等级和参考值可设置如表 1 所示，解读如下：

社交账户其使用年限可分为“短(S)”、“中(M)”、“长(L)”三个参考等级。根据经验，可设置年限“短”的参考值为 0 年。同时，年限“中”和“长”的参考值分别设置为 4 年和 10 年。

对于账户的认证情况，认证的数量越多其可信度越高。账户认证数的参考等级可设置为“无(NO)”、“少(S)”和“多(M)”，其参考值可分别设置为 0 项、2 项和 5 项。

账户粉丝数的参考等级可设置为四个，即“无(NO)”、“少(S)”、“一般(N)”和“多(M)”。四个参考等级分别设置参考值为 0 个、3000 个、5000 个和 15000 个。

被转赞评数的参考等级设置为“无(NO)”、“少(S)”和“多(M)”，参考值设置为 0 个、50000 个和 150000 个。

信息原创率和存疑率的参考等级可设置为“低(L)”、“中(M)”和“高(H)”，相应参考值设置为 0%、50%和 100%。

表 1　指标参考等级和参考值

Tab. 1　The referential levels and referential values

C_1		C_2		C_3		C_4		C_5		C_6	
等级	值	等级	值	等级	值	等级	值	等级	值	等级	值
S	0	NO	0	NO	0	NO	0	L	0	L	0
M	4	S	2	S	3000	S	50000	M	0.5	M	0.5
L	10	M	5	N	5000	M	150000	H	1	H	1
—	—	—	—	M	15000	—	—	—	—	—	—

结合上述指标参考等级和参考值，通过专家经验及数据的趋势性分析，可以针对各子模型的规则库给出初始参数，如表 2 所示。IHBRB 中 4 个子模型输入的初始权重均设置为 1，规则的初始权重均设置为 1。此外，模型中 6 个输入指标转换函数的初始自适应系数均设置为 1，即 $s_1=s_2=s_3=s_4=s_5=s_6=1$。在此情况下，该转换函数为线性函数。

5.3　模型优化与结果分析

由于专家认知的局限性，需要结合有限的数据样本对 IHBRB 模型的初始参数进行调整。优化目标函数依据式(12)和式(13)确定。

在所获取的数据集中，随机取 50%的数据作为训练集，以整个数据集作为测试集。对于 P-CMA-ES 优化算法，初始参数向量 K^0 即为初始模型参数、初始协方差矩阵 C^0 为单位阵，初始步幅设置为 $\rho^0=0.5$，优化迭代次数设置为 200 代。

优化后的 IHBRB 模型的规则权重和置信度如表 3 所示，其中优化后的各子模型的指标权重分别为：0.75、1、0.55、1、0.3、1、0.93、0.91 和 1。优化后转换函数的自适应系数分别为：1.61、0.35、0.59、1.72、2.18 和 1.83。采用初始 IHBRB 和优化后的 IHBRB 模型对测试集中账户进行可信度评估，评估结果如图 7 所示，可见优化后的模型能够更好实现对社交账户可信度的评估。初始模型和优化后模型的输出结果均方误差分别为 0.0562 和 0.0028，优化后模型的精度提高了 95%。

5.4　对比研究

为了进一步验证所提方法的有效性，在本节的对比研究中，采用无自适应系数转换函数的 IHBRB 模型(记为 IHBRB-1)、神经网络模型(记为 BPNN)、模糊推理模型(记为 FRM)和回归支持向量机模型(记为 SVR)对前述数据集中的账户进行可信度评估。BPNN、FRM、SVR 为三种常用的评估模型。其中，BPNN 模型具有精度高、易操作等优点；FRM 模型能够有效描述和处理模糊不确定性，且能够较好融合专家判断；SVR 模型不依赖于优化数据样本的数量，具有较高的建模精度。对比实

验中，随机选取 50%的数据作为训练集，以整个数据集为测试集，进行 10 轮重复实验，以结果的平均值为最终结果。

表 2　初始 IHBRB 模型

Tab. 2　The initial IHBRB

BRB-1	C_1	C_2	规则结论	BRB-4	B_1	B_2	B_3	规则结论
1	S	NO	{0.5, 0.5, 0}	1	U	U	U	{0.6, 0.3, 0.1}
2	S	S	{0.3, 0.6, 0.1}	2	U	U	P	{0.4, 0.6, 0}
3	S	M	{0, 0.8, 0.2}	3	U	U	F	{0.1, 0.2, 0.7}
4	M	NO	{0.4, 0.5, 0.1}	4	U	P	U	{0.5, 0.3, 0.2}
5	M	S	{0.1, 0.1, 0.8}	5	U	P	P	{0, 0.6, 0.4}
6	M	M	{0.1, 0.2, 0.7}	6	U	P	F	{0.1, 0.4, 0.5}
7	L	NO	{0, 0.2, 0.8}	7	U	F	U	{0.2, 0.2, 0.6}
8	L	S	{0.3, 0.6, 0.1}	8	U	F	P	{0.2, 0.2, 0.6}
9	L	M	{0.1, 0.3, 0.6}	9	U	F	F	{0, 0.7, 0.3}
BRB-2	C_3	C_4	规则结论	10	P	U	U	{0, 0.6, 0.4}
1	NO	NO	{0.1, 0.7, 0.2}	11	P	U	P	{0.6, 0.3, 0.1}
2	NO	S	{0.2, 0.7, 0.1}	12	P	U	F	{0.1, 0.7, 0.2}
3	NO	M	{0.2, 0.5, 0.3}	13	P	P	U	{1, 0, 0}
4	S	NO	{0, 0.9, 0.1}	14	P	P	P	{0.9, 0.1, 0}
5	S	S	{0, 0.8, 0.2}	15	P	P	F	{0.3, 0.6, 0.1}
6	S	M	{0.1, 0.8, 0.1}	16	P	F	U	{0.1, 0.4, 0.5}
7	N	NO	{0.4, 0.5, 0.1}	17	P	F	P	{0, 0, 1}
8	N	S	{0.3, 0.7, 0}	18	P	F	F	{0.1, 0.3, 0.6}
9	N	M	{0, 0.2, 0.8}	19	F	U	U	{0.1, 0.6, 0.3}
10	M	NO	{0.3, 0.5, 0.2}	20	F	U	P	{0.6, 0.3, 0.1}
11	M	S	{0.1, 0.6, 0.3}	21	F	U	F	{0.3, 0.5, 0.2}
12	M	M	{0, 0.1, 0.9}	22	F	P	U	{0.1, 0.1, 0.8}
BRB-3	C_5	C_6	规则结论	23	F	P	P	{0.2, 0.5, 0.3}
1	L	L	{0, 0.2, 0.8}	24	F	P	F	{0.3, 0.6, 0.1}
2	L	M	{0.8, 0.1, 0.1}	25	F	F	U	{0.6, 0.3, 0.1}
3	L	H	{0.4, 0.5, 0.1}	26	F	F	P	{0.3, 0.6, 0.1}
4	M	L	{0, 0.4, 0.6}	27	F	F	F	{0.8, 0.1, 0.1}
5	M	M	{0.2, 0.8, 0}					
6	M	H	{0.4, 0.6, 0}					
7	H	L	{0.1, 0.6, 0.3}					
8	H	M	{0.5, 0.3, 0.2}					
9	H	H	{0.2, 0.5, 0.3}					

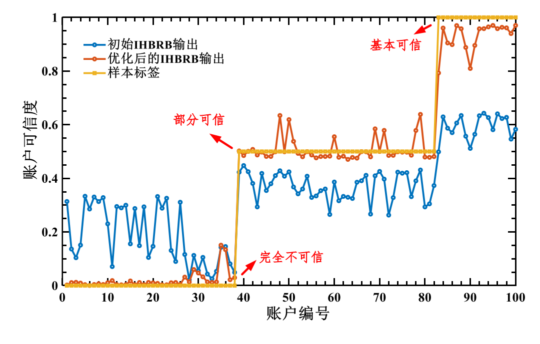


图 7　账户可信度评估结果

Fig. 7　The results of account credibility evaluation

实验结果如图 8 所示，可以看出这些模型经过优化均能够从一定程度上反映账户的可信度。对比 IHBRB 与 IHBRB-1 结果可以看出，转换函数自适应系数的引入对评估精度的提高有较大作用。FRM 对“完全不可信”的评估精度高于对“部分可信”和“基本可信”的评估精度。

对于上图中的某一账户样本，其账户年限为 10 年，认证数为 5 项，粉丝数为 37422 人，被转赞评数为 219054 人次，信息原创率为 97%以及信息存疑率为 3%，通过专家判断其为基本可信账户。采用上述 5 种模型评估该账户：IHBRB 输出的置信分布为{0.00, 0.16, 0.84}，对应可信度效用为 0.92；IHBRB-1 输出的置信分布为{0.05, 0.21, 0.74}，对应可信度效用为 0.84；BPNN 输出的可信度为 0.81；FRM 输出中对各可信度等级的隶属度分别为 0.08, 0.33 和 0.59，对应的可信度效用为 0.755；SVR 输出的可信度值为 0.79。可见，这些模型均认为该账户属于“基本可信”，但是 IHBRB 模型不仅能够输出相对于各等级的置信度，其量化评估的可信度更接近真实值。

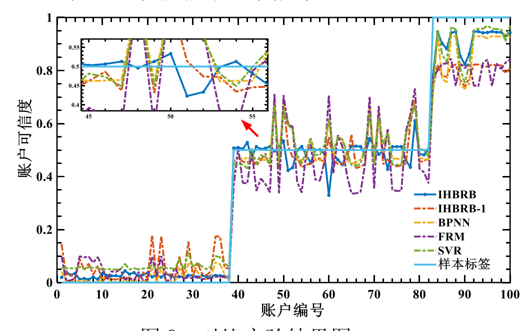


图 8　对比实验结果图

Fig. 8　The result of comparative study

表 3　优化的 IHBRB 模型

Tab. 3　The optimized IHBRB

BRB-1	权重	规则结论	BRB-4	权重	规则结论
1	0.60	{0.85, 0.09, 0.06}	1	0.73	{0.23, 0.7, 0.07}
2	0.31	{0.36, 0.44, 0.2}	2	0.55	{0.31, 0.2, 0.49}
3	0.09	{0.49, 0.15, 0.36}	3	0.95	{0.34, 0.27, 0.39}
4	0.51	{0.73, 0.14, 0.13}	4	0.24	{0.47, 0.23, 0.3}
5	0.21	{0.45, 0.06, 0.49}	5	0.75	{1, 0, 0}
6	0.28	{0.52, 0.14, 0.34}	6	0.06	{0.36, 0.29, 0.35}
7	0.30	{0.08, 0.18, 0.74}	7	0.72	{0.32, 0.58, 0.1}
8	0.24	{0.54, 0.16, 0.3}	8	0.24	{0.56, 0.27, 0.17}
9	0.46	{0.03, 0.13, 0.84}	9	0.34	{0.33, 0.31, 0.37}
BRB-2	权重	规则结论	10	0.48	{0.14, 0.41, 0.45}
1	0.22	{0.13, 0.75, 0.12}	11	0.21	{0.2, 0.53, 0.27}
2	0.09	{0.33, 0.1, 0.57}	12	0.50	{0.38, 0.45, 0.17}
3	0.30	{0.25, 0.49, 0.26}	13	0.48	{0.38, 0.39, 0.23}
4	0.61	{0.11, 0.76, 0.13}	14	0.07	{0.31, 0.38, 0.31}
5	0.65	{0.16, 0.13, 0.71}	15	0.58	{0.32, 0.51, 0.17}
6	0.21	{0.17, 0.31, 0.52}	16	0.36	{0.37, 0.29, 0.34}
7	0.14	{0.26, 0.33, 0.41}	17	0.42	{0.32, 0.35, 0.33}
8	0.21	{0.11, 0.61, 0.28}	18	0.75	{0.4, 0.27, 0.33}
9	0.30	{0.27, 0.57, 0.16}	19	0.58	{0.46, 0.33, 0.21}
10	0.10	{0.26, 0.16, 0.58}	20	0.53	{0.41, 0.21, 0.38}
11	0.74	{0, 0, 1}	21	0.71	{0.48, 0.35, 0.17}
12	0.89	{0.31, 0.51, 0.18}	22	0.05	{0.25, 0.32, 0.43}
BRB-3	权重	规则结论	23	0.83	{0.03, 0, 0.97}
1	1.00	{0.12, 0.48, 0.4}	24	0.40	{0.23, 0.39, 0.38}
2	0.74	{0.39, 0.45, 0.16}	25	0.65	{0.62, 0.06, 0.32}
3	0.64	{0.26, 0.33, 0.41}	26	0.88	{0.49, 0.22, 0.29}
4	0.32	{0.29, 0.52, 0.19}	27	0.31	{0.54, 0.11, 0.35}
5	0.08	{0.21, 0.1, 0.69}			
6	0.35	{0.01, 0.46, 0.53}			
7	0.07	{0.35, 0.39, 0.26}			
8	0.39	{0.4, 0.07, 0.53}			
9	0.89	{0.48, 0.48, 0.04}			

为了研究训练样本的减少对实验结果的影响，分别随机选择 40%、30%和 20%的数据样本作为训练集。实验结果如表 4 所示，可以看出，当训练集的数量下降，5 种模型输出的均方误差都增大。但是对于 IHBRB、IHBRB-1 和 FRM 这类能够通过专家知识确定初始参数的模型，其在优化数据样

本较少时，仍能够达到较好评估性能。

表 4 对比模型的均方误差

Tab. 4 The mean square error of the comparative model

数据样本	IHBRB	IHBRB-1	BPNN	FRM	SVR
50%	0.0028	0.0089	0.0074	0.0121	0.0087
40%	0.0044	0.0097	0.0111	0.0152	0.0219
30%	0.0102	0.0174	0.0453	0.0341	0.0476
20%	0.0236	0.0303	0.0951	0.0613	0.0899

为了进一步说明 P-CMA-ES 算法的有效性，分别采用 PSO 算法和 DE 算法对初始 IHBRB 模型进行优化。数据集和重复实验次数均与前述一致，优化后模型输出的均方误差如表 5 所示。可见，当训练集比例为 50%时，三种方法优化算法的性能相近。当训练集比例降低为 20%时，通过 P-CMA-ES 算法优化的 IHBRB 模型具有更好的建模精度。

表 5 不同优化算法下 IHBRB 输出的均方误差

Tab. 5 The mean square error of IHBRB with different optimization algorithms

数据样本	P-CMA-ES	PSO	DE
50%	0.0028	0.0030	0.0031
40%	0.0044	0.0077	0.0089
30%	0.0102	0.0147	0.0166
20%	0.0236	0.0281	0.0273

6 结束语

本文针对社交账户可信度评估面临的指标多维且特性各异、数据信息不确定性多样等问题，从账户属性、交际属性和内容属性三个角度构建了账户可信度评估指标体系，提出了一种基于改进分层 BRB 的可信度评估方法。该方法通过分层结构和置信框架来描述指标间的关系和信息不确定性，同时引入了带有自适应系数的信息转换函数以更好处理指标间的特性差异。实验证明，该方法能够准确评估社交账户的可信度。但本文仍存在一些不足需要进一步研究，如可与深度学习等技术融合，挖掘更多信息以构建规则库，进一步提高评估过程的可解释性及输出结果的精度。

参考文献：

[1] 胡学韬，陈秀真. 基于信任度评估的社交网络虚假账户检测 [J]. 信息安全与通信保密, 2014, (5): 90-94. (Hu Xuetao, Chen Xiuzhen. Fake account detection in online social network based on trust evaluation [J]. Information Security and Communications Privacy, 2014, (5): 90-94.)

[2] 路金泉. 面向社交网络信息管控的信息可信度评估方法研究 [D]. 郑州：解放军信息工程大学, 2017. (Lu Jinquan. Research on information credibility evaluation method for social network information control [D]. Zhengzhou: PLA Information Engineering University, 2017.)

[3] 王坤. 在线社交网络异常账户检测算法研究 [D]. 西安：西安电子科技大学, 2020. (Wang Kun. Research on online social network abnormal account detection algorithm [D]. Xi’an: Xidian University, 2020)

[4] 刘亚尚. 在线社交网络异常账户检测技术研究 [D]. 南京师范大学, 2016. (Liu Yashang. Research on anomaly detection methods for online social networks [D]. Nanjing: Nanjing Noraml University, 2016.)

[5] 王峥，叶维，邱秀连. 基于特征加权贝叶斯神经网络的微博异常账号检测 [J]. 计算机与数字工程, 2018, 46 (11): 2323-2328. (Wang Zheng, Ye Wei, Qiu Xiulian. Weibo abnormal account detect based on weighted bayesian neural network [J]. Computer and Digital Engineering, 2018, 46 (11): 2323-2328)

[6] Yang Jianbo, Liu Jun, Wang Jin, *et al.* Belief rule-base inference methodology using the evidential reasoning approach-RIMER [J]. IEEE Trans on Systems Man and Cybernetics: Systems, 2006, 36 (2): 266-285.

[7] 周志杰，杨剑波，胡昌华，等. 置信规则库专家系统与复杂系统建模 [M]. 北京：科学出版社, 2011. (Zhou Zhijie, Yang Jianbo, Hu Changhua, *et al.* Belief rule base expert systems and complex system modeling [M]. Beijing: Science Press, 2011.)

[8] 周志杰，曹友，胡昌华，等. 基于规则的建模方法的可解释性及其发展 [J]. 自动化学报, 2021, 47 (06): 1201-1216. (Zhou Zhijie, Cao You, Hu Changhua, *et al.* The interpretability of rule-based modeling approach and its development. Acta Automatica Sinica, 2021, 47 (06): 1201-1216.)

[9] 王越，张剑金，刘芳芳. 一种多特征微博僵尸粉检测方法与实现 [J]. 中国科技论文, 2014, 9 (1): 81-86. (Wang Yue, Zhang Jianjin, Liu Fangfang. Detection of micro-blog zombie fans based on multi-features [J]. China Sciencepaper, 2014, 9 (1): 81-86.)

[10] 王培人，毛剑，马寒军等. 基于用户信息的社交网络信任评估方法 [J]. 计算机应用研究, 2018, 35 (2): 521-526. (Wang Peiren, Mao Jian, Ma Hanjun, *et al.* User information based trust evaluation mechanism for social network [J]. Application Research of Computers, 2018, 35 (2): 521-526.)

[11] 王路帮，李守伟. 社交网络中基于网络结构支撑模型的谣言传播研究 [J]. 计算机应用研究, 2019, 36 (10): 3094-3097. (Wang Lubang, Li Shouwei. Rumor spreading based on network structural supportiveness model in social network [J]. Application Research of Computers, 2019, 36 (10): 3094-3097.)

[12] Yang Jianbo, Liu Jun, Xu Donglin, *et al.* Optimization models for training belief-rule-based systems [J]. IEEE Trans on Systems Man and Cybernetics: Systems, 2007, 37 (4): 569-585.

[13] Zhou Zhijie, Hu Guanyu, Hu Changhua, *et al.* A survey of belief rule-base expert system [J]. IEEE Trans on Systems Man and Cybernetics: Systems, 2021, 51 (8): 4944-4958.

[14] 胡蓉，易照云，钱斌. 基于置信规则库的油浸式变压器故障诊断 [J]. 北京工业大学学报, 2021, 47 (9): 1000-1010. (Hu Rong, Yi Zhaoyun, Qian Bin. Fault Diagnosis of Oil-immersed Transformer Based on Belief Rule Base [J]. Journal of Beijing University of Technology, 2021, 47 (9): 1000-1010.)

[15] 胡冠宇. 基于置信规则库的网络安全态势感知技术研究 [D]. 哈尔滨：哈尔滨理工大学, 2016. (Hu Guanyu. Study on network security situation awareness based on belief rule base [D]. Harbin: Harbin University of Technology, 2016.)

[16] 刘栅杉，朱海龙，韩晓霞，等. 基于主成分回归和分层置信规则库的企业风险评估模型 [J]. 计算机科学, 2021, 48 (Z2): 570-575. (Liu Shanshan, Zhu Hailong, Han Xiaoxia, *et al.* Enterprise risk assessment model based on principal component regression and hierarchical belief rule base [J]. Computer Science, 2021, 48 (Z2): 570-575.)

[17] Chang Leilei, Zhou Zhijie, Liao Huchang, Generic disjunctive belief rule base modeling, inferencing, and optimization [J]. IEEE Trans on Fuzzy Systems, 2019, 27 (9): 1866-1880.

[18] Qian Bin, Wang Qianqian, Hu Rong, *et al.* An effective soft computing technology based on belief-rule-base and particle swarm optimization for tipping paper permeability measurement [J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10 (3): 841-850.